

## STATISTIEK MET DECEMBERZEGELS

Piet Jonkman



Vroeger hadden we de opklimmende reeks “leugens, kranten en statistiek”. In deze tijd zullen we dit rijtje kunnen uitbreiden met AI als overtreffende trap van onwaarheid. En misschien moeten we nu de (papieren) krant uit dit rijtje verwijderen als wellicht nog meest betrouwbare bron voor ware berichtgeving.

Dat statistiek als zo onbetrouwbaar werd (wordt?) beschouwd heeft misschien te maken met de zwaarte van het vak. In mijn opleiding gebeurde het vaak dat studenten meerdere keren opgingen voor het vak totdat ze uiteindelijk de statistische kans op een voldoende resultaat bereikten. Ik ken ook een Amsterdamse student die geheel zijn studententijd moedig pogde dit vak te halen en daarin vlak voor zijn afrondend examen slaagde.

Feit is wel dat het een ondoordringbaar vak is dat met behulp van ingewikkelde wiskundige formules aantoont wat met gezond verstand ook te beredeneren valt. Ook kun je met dit vak alles bewijzen. En vooral als de uitkomst niet welgevallig is zijn er veel mogelijkheden om statistisch aan te tonen dat de uitgangspunten niet goed zijn of dat er iets aan de steekproef mankeert.

Genoeg van deze gefrustreerde blik op het verleden. Het schoot me weer te binnen bij mijn beschouwing van de decemberzegels.

Dat je niet van allemaal evenveel ontvangt is te begrijpen. Maar als je veel zegels hebt dan mag je toch verwachten dat de aantallen naar elkaar toekruipen.

Niets is minder waar in mijn geval. Ik krijg van alle kanten zegels, dus mijn waarneming is niet georganiseerd. En ook de post die ik zelf ontving komt van meerdere kanten, dus ik durf te stellen dat mijn steekproef/waarneming onafhankelijk is. Hooguit kun je stellen dat deze te klein is om daaraan conclusies te verbinden.

We nummeren de zegels van links boven (1) naar rechtsonder (10). De scores zijn van midden maart. 2024.

Van zegel **1** heb ik 16 exemplaren, van **2** 15, enzovoort.

Decemberzegels 2023				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
16	15	11	13	10
<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
2	21	5	7	11

We hebben dus in totaal 111 zegels. Het gemiddelde is dus (ongeveer zegt de ingenieur) 11. De aantallen van de meeste zegels komen daar in de buurt. Uitschieter naar boven is zegel **7**, terwijl de zegels **6**, **8** en **9** sterk naar beneden afwijken. Tot zover de waarnemingen.

Nu komt het meest interessante deel, namelijk de verklaring.

Je kunt aannemen dat de plaats van de zegel in het velletje mede bepalend is voor het gebruik. Zegel **1**, **5**, **6** en **10** zouden dan het hoogst scoren daar vrijwel ieder geneigd is van buiten naar binnen te werken. Deze aanname kan ik zelf logenstraffen daar op mijn restantvelletje alleen de zegels **6** en **7** zijn overgebleven. Merkwaardigerwijs juist die zegels die in de tabel de grootste afwijking van het gemiddelde vertonen.

Een andere aanname is dat mensen bewust een plaatje uitkiezen. Dat was mijn eerste conclusie want ik had al heel veel “uiltjes”(4) terwijl ik van de andere zegels nog geen of nog maar een paar had. Maar later in de tijd werd die ruim overklast door de zegel van het Noorderlicht (**7**). En eigenlijk kan ik me ook niet voorstellen dat dit een goede veronderstelling is, want je kunt het per velletje maar één keer doen en bovendien is het niet logisch, daar de meeste mensen toch wel systematisch werken (aanname 1).

Het meest intrigerend vind ik dat de zegel links onder (**6**) zo weinig voorkomt. Dat strijdt met aanname 1 en zo veel slechter dan de andere zegels (aanname 2) is deze toch ook niet?

Wel is het weer logisch dat de zegels van de bovenste rij hoger scoren dan de zegels van de onderste rij. De meeste mensen zullen gewoon links bovenaan beginnen en dan is het voor de hand liggend dat als er zegels overblijven dat zegels van de onderste rij zijn. Maar weer blijft dan de vraag hoe die uitschieter van zegel 7 is te verklaren.

Wie weet nog een aanname te bedenken om de verschillen te verklaren? Of blijft de conclusie “statistiek is een ondoorgrondelijk vak”?